

## USING DATA FROM OTHER SOURCES — A CHECKLIST FOR QUALITY CONCERNS

1. Identify the decision you are making or project objectives (a standard part of project planning or hypothesis testing).
2. Identify the data and information from outside sources proposed for the project/decision support. Note that this may not be obvious. Include data bases, maps and literature, and don't overlook:
  - Information and data used to site or time sampling events (meteorology, geology, etc.)
  - Anecdotal or other information triggering the study
  - Toxicity, exposure, and environmental fate data
  - Models and their output
  - Census data
  - GIS data
3. To avoid investigating information that may not be suitable, determine whether these data have any non-quality constraints affecting their use in the new project/decision support. That is, are there programmatic, legal, or other constraints on the use of the data? Example, is it proprietary or CBI? Does your new project/decision have programmatic constraints requiring only "approved" sources, required peer review or validation of draft data? Obviously, if your proposed data fail these checks, you may not be able to use them, and need not continue down this checklist. If you still plan to use them, you must modify your expectations about the applicability of the project/decision. Check with the program involved. Here are examples:
  - CAA Credible Evidence Revisions (FR 62:36, Feb. 24, 1997)
  - Federal Rule of Evidence 702
4. If not incidental to step 2, determine where the acquired data will be used in the decision making process. That is, will it be used to scope the new project, contribute to data collection in the project, verify the results of the decision, substitute for all or some new data collection, etc.?

Case 1: If acquired data will be the basis of comparison for new data, the former's quality should be investigated first, before the new effort begins. This is to both ensure that it is worth the effort of further study and prevent "apples and oranges" results. An example of what can go wrong when this is not done was in the news when Tulane University withdrew a peer reviewed paper published in Science on the effects of mixtures of estrogen. EPA had already mobilized research efforts based on the results. The researcher (and others) found that

the results could not be replicated and a study design flaw was suspected (“Tulane University Withdraws Paper That Prompted Health Fears”, NY Times, AP, 8/21/97). This does not imply that all results must be replicated before use, but a consideration of the decision being made as a result of a single study is warranted.

- Case 2: If the data are part of scoping or design for more data collection (for example, a pilot project, background historical data, or sample survey), many quality issues can be determined by the resolution of the new effort. If ballpark estimates are good enough, proven reliability of acquired data may be unnecessary. This is especially true if new sampling alone will lead to the decision. Data usability in this case is an individual matter, to be determined by the project manager with statistical help.
- Case 3: If the acquired data or information are not directly used to compute results, they will still affect the results. An example is the use of existing locational, geological, hydrological, or meteorology data used to locate or time sample collection. The materials and methods involved in producing these data are one consideration, but the quality assurance system implemented to ensure the results were reliable is also important. The source of this information is frequently public domain and used without question. Beware of assuming it is ok especially if it is critical to the new project/decision. How will it affect the outcome?
- Case 4: If the acquired data will totally substitute for any new data collection efforts, a comprehensive analysis of the past quality assurance controls and hypothetical needs may be required. The effect of the data quality on the decision will directly affect the intensity of effort to determine and document the quality of the data. This sounds like circular reasoning, but the scrutiny of the data will need to match the importance of the decision based upon it, and its contribution to the decision. See the annotated references after item 6. The best guidance for this effort is G-4, if the data quality objectives need to be iteratively applied; R-5, if the QAPP needs to be reconstructed, and G-9, for assessing the data in light of the study objectives.
- Case 5: A variation of Case 4's substitution for new data collection efforts is a partial substitution of acquired data for new data, for example, in modeling and risk assessment. Some parameters, like environmental concentration data are newly collected, but modeling and other data are

used to infer concentrations in other media, at other times, and in people, animals and plants. Quality concerns can be spread in so many directions, that sorting out the crucial ones seems overwhelming. It has been attempted, however, because the consequences of error are frequently staggering in terms of cost and the health of humans and the environment. See the risk assessment-related references under item 6.

5. Scrutinize data/information for quality concerns pertinent to the intended use. The most straightforward way to approach data quality is to retroactively apply your new data collection standards to the data. For each procedure that was or would have been documented in its QAPP, ask whether it is known and acceptable for the intended use if known. If unknown, first, is it important to the new project/decision, and second, can it be discovered or inferred? This implies the use of the “graded” approach. The ultimate set of quality standards for judging the data are those dictated by the intended use.

For Cases 1, 2, 4, 5: Begin by applying data quality objectives, or discerning those of the existing study (for Case 1). EPA QA/G-9, Box 1.1-1: Example Applying the DQO Process Retrospectively (1 page).

***Note:** Some programs already document decisions based on data from specified outside sources. The decision to use the data is specified by direct comparison with program criteria for their acceptance. The program’s decision to use outside sources is presumably made based on their DQOs, documented as such and kept for the public record. The concerns over the data quality are therefore relatively specific given the same source and same type of decision. For example, if public utility data is always the source of emissions inventory data used in setting emissions standards, the program should have DQO-derived acceptance criteria. A valid question is whether all standards can be set with comparable quality if there are fewer data points for one as opposed to many for another. A statistician can help to answer this question for any particular data set. Examples of these programs within EPA include ETV, some air and solid waste rule making and standards, etc. The program QA managers can direct users to these DQOs and data acceptance criteria, and possibly model QAPPs for their use.*

For Case 3: For instances where data/information are used in a project or decision that are not the quantitative result per se, some critical thinking is involved. Qualitative information can not be compared readily to DQOs, but their effect on the outcome should be examined. If it is important to the decision, a justification for using it should be supplied. For quantitative data derived in other studies yet important to designing, sampling, or modeling results, the quality

should be noted if a “devil’s advocate” approach indicates that problems with its quality could alter project outcomes/decisions. Many of these concerns are routinely left for study users to consider by simply noting the source of the information and possibly how or why it was used. If a future user of the information has a problem with NOAA’s climate data, they know the study used it. If less familiar sources of information are used, for example a local hydrogeological study for placement of samples, or local reports on presence of certain species key to sampling locations, the reasons for accepting their quality should be sought and noted if it is key to the project/decision. For investigations of certain quality aspects of acquired information, see the guidance for the other cases.

*Example for comparing information to intended use:* If the acquired data represent historical pollutant loads in a water body measured in the spring and the decision to be made must address year-round loading, this must be acknowledged as a factor biasing the decision, if the decision can be made at all. Whether the data were collected with adequate QA oversight, acceptable methods, by trained samplers, and analyzed with proper holding times, accurate methods with acceptable detection limits, may be moot. If temporal concerns are not important to the decision, but the absolute concentration will trigger a decision of great importance, a thorough examination of the QA and QC practices by the data collectors would be essential.

6. Document your analysis plan in a QAPP. If the project also includes some new data collection, list and indicate your intention to investigate the acquired data based on anticipated effects upon the results of that effort. The acquired data investigation results may be a determining step in a decision to proceed with data collection. If the acquired data is substituted for any new data collection efforts, a QAPP is still required. Remember that the graded approach does apply, and many sections of the QAPP will not apply if there is no new data collection. Depending upon the nature of the decision being made, very little may be required. Turn to QA staff, QA and other guidance for specific assistance in documenting your use of acquired data.

Remember that the original data/information collector(s) may also be the best source of information on the quality system under which it was collected. This information, also called “meta data” may not be published, but it may yet exist. Unfortunately, even if QA requirements exist, it may be difficult to determine if they were followed. For example, even the original sampling and analysis plan or QAPP may not have been followed as written.

7. Execute your analyses and document the outcome appropriately (for the program’s graded approach) relevant to the decision or project.

## Useful References

The following sources may be useful in the development of acceptance criteria/limitations for the use of data collected for other purposes in order to ensure that it is adequate for the new purpose. This acceptance criteria is documented in a Quality Assurance Project Plan. Note that in some EPA programs, a legislative mandate can determine how the Agency uses data from outside. Examples include the CAA Credible Evidence Revisions (FR 62:36, Feb. 24, 1997) and Federal Rule of Evidence 702.

Chalmers, I.; Altman, D.O., eds. *Systematic Reviews* (London, United Kingdom: BMJ Publishing Group 1995)

Ducharme, M.K.; Licklider, B.L.; Matthes, W.A.; Vannatta, R.A., *Conceptual and Analysis Criteria: A Process for Identifying Quality Educational Research* (Des Moines, IA: FINE Foundation 2000)

Fink, Arlene, *Conducting Research Literature Reviews: from paper to the Internet* (Thousand Oaks, CA: Sage Publications, Inc. 1998)

Greenhalgh, T., *How to Read a Paper: The Basics of Evidence Based Medicine* (London, United Kingdom: BMJ Publishing Group 1997)

Exposure Factors Handbook, Volume I of III - General Factors, Update to Exposure Factors Handbook, EPA/600/8-89/043 - May 1989 (EPA/600/P-95/002Ba, August 1996, SAB Review Draft).

Guidance for Data Useability in Risk Assessment (Part A & B), April & May 1992, 9285.7-09A&B, Superfund.

Kollig, H.P. 1988. Criteria for Evaluation the Reliability of Literature Data on Environmental Process Constants. *Toxicological and Environmental Chemistry* 17(4):287-311.

Kollig, H.P. 1990. A Fate constant Data Program. *Toxicological and Environmental Chemistry* 25(2-3):171-179.

Kollig, H.P. and B.E. Kitchens, "Problems Associated with Published Environmental Fate Data". *Toxicological and Environmental Chemistry* 28(2-3):95-103.

Kollig, Heinz. 1993. Environmental Fate Constants for Organic Chemicals Under consideration for EPA's Hazardous Waste Identification Projects. U.S. EPA, Athens, GA, EPA/600/R-93/132.

Kollig, H.P. 1995. Environmental Fate Constants for Additional 27 Organic Chemicals Under Consideration for EPA's Hazardous Waste Identification Projects. U.S. EPA, Athens, GA, EPA/600/R-95/039.

Nolan, Melvin, "ORD-NCEA Guidelines for Developing QAPPs for Research Projects Using Existing Peer Reviewed Literature and Meta-data, and Secondary Data", ORD/National Center for Environmental Assessment.

Schumacher, B.A. and Conkling, B. L. User's Guide to the QA/QC Evaluation Scale of Historical Datasets (17p), December 20,1990, unpublished manuscript. First author currently with EPA in Las Vegas.